

## PATENT COOPERATION TREATY

REC'D 30 SEP 2005

PCT

WIPO

PCT

## INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 800293WO	<b>FOR FURTHER ACTION</b>	
	See Form PCT/PEA/416	
International application No. PCT/FI2004/000331	International filing date (day/month/year) 01.06.2004	Priority date (day/month/year) 06.06.2003
International Patent Classification (IPC) or national classification and IPC G06F17/30		
Applicant TIETOENATOR OYJ		
<p>1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.</p> <p>2. This REPORT consists of a total of 6 sheets, including this cover sheet.</p> <p>3. This report is also accompanied by ANNEXES, comprising:</p> <p>a. <input checked="" type="checkbox"/> (<i>sent to the applicant and to the International Bureau</i>) a total of 5 sheets, as follows:</p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> sheets of the description, claims and/or drawings which have been amended and are the basis of this report and/or sheets containing rectifications authorized by this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions).</li> <li><input type="checkbox"/> sheets which supersede earlier sheets, but which this Authority considers contain an amendment that goes beyond the disclosure in the international application as filed, as indicated in item 4 of Box No. I and the Supplemental Box.</li> </ul> <p>b. <input type="checkbox"/> (<i>sent to the International Bureau only</i>) a total of (indicate type and number of electronic carrier(s)), containing a sequence listing and/or tables related thereto, in computer readable form only, as indicated in the Supplemental Box Relating to Sequence Listing (see Section 802 of the Administrative Instructions).</p>		
<p>4. This report contains indications relating to the following items:</p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Box No. I Basis of the opinion</li> <li><input checked="" type="checkbox"/> Box No. II Priority</li> <li><input type="checkbox"/> Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability</li> <li><input type="checkbox"/> Box No. IV Lack of unity of invention</li> <li><input checked="" type="checkbox"/> Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement</li> <li><input type="checkbox"/> Box No. VI Certain documents cited</li> <li><input type="checkbox"/> Box No. VII Certain defects in the international application</li> <li><input type="checkbox"/> Box No. VIII Certain observations on the international application</li> </ul>		
Date of submission of the demand 05.04.2005	Date of completion of this report 04.10.2005	
Name and mailing address of the International preliminary examining authority:  European Patent Office - P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk - Pays Bas Tel. +31 70 340 - 2040 Tx: 31 651 epo nl Fax: +31 70 340 - 3016	Authorized Officer Warry, L Telephone No. +31 70 340-3199	



**INTERNATIONAL PRELIMINARY REPORT  
ON PATENTABILITY**

International application No.  
PCT/FI2004/000331

**Box No. I Basis of the report**

1. With regard to the **language**, this report is based on the international application in the language in which it was filed, unless otherwise indicated under this item.
  - This report is based on translations from the original language into the following language, which is the language of a translation furnished for the purposes of:
    - international search (under Rules 12.3 and 23.1(b))
    - publication of the international application (under Rule 12.4)
    - international preliminary examination (under Rules 55.2 and/or 55.3)
2. With regard to the **elements\*** of the international application, this report is based on (*replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report*):

**Description, Pages**

1-22 as originally filed

**Claims, Numbers**

1-28 received on 21.07.2005 with letter of 19.07.2005

**Drawings, Sheets**

1-8 as originally filed

a sequence listing and/or any related table(s) - see Supplemental Box Relating to Sequence Listing

- The amendments have resulted in the cancellation of:
  - the description, pages
  - the claims, Nos.
  - the drawings, sheets/figs
  - the sequence listing (*specify*):
  - any table(s) related to sequence listing (*specify*):
- This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).
  - the description, pages
  - the claims, Nos.
  - the drawings, sheets/figs
  - the sequence listing (*specify*):
  - any table(s) related to sequence listing (*specify*):

\* If item 4 applies, some or all of these sheets may be marked "superseded."

**INTERNATIONAL PRELIMINARY REPORT  
ON PATENTABILITY**

International application No.  
PCT/FI2004/000331

**Box No. II Priority**

1.  This report has been established as if no priority had been claimed due to the failure to furnish within the prescribed time limit the requested:
  - copy of the earlier application whose priority has been claimed (Rule 66.7(a)).
  - translation of the earlier application whose priority has been claimed (Rule 66.7(b)).
2.  This report has been established as if no priority had been claimed due to the fact that the priority claim has been found invalid (Rule 64.1). Thus for the purposes of this report, the international filing date indicated above is considered to be the relevant date.
3. Additional observations, if necessary:

**Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

**1. Statement**

Novelty (N)	Yes: Claims	
	No: Claims	1-28
Inventive step (IS)	Yes: Claims	
	No: Claims	1-28
Industrial applicability (IA)	Yes: Claims	1-28
	No: Claims	

**2. Citations and explanations (Rule 70.7):**

**see separate sheet**

**INTERNATIONAL PRELIMINARY  
REPORT ON PATENTABILITY  
(SEPARATE SHEET)**

International application No.  
PCT/FI2004/000331

**Re Item V.**

**1.0 The following document is referred to in this communication:**

D1: SA LIN ET AL: "Integrating a heterogeneous distributed data environment with a database specific ontology" PROCEEDINGS OF THE ISCA 14TH INTERNATIONAL CONFERENCE PARALLEL AND DISTRIBUTED COMPUTING SYSTEMS INT. SOC. COMPUT. & THEIR APPLICATIONS - ISCA CARY, NC, USA, 2001, pages 430-435, XP002297454 ISBN: 1-880843-39-0

1.1 Claims 1-28 are not novel (Article 33(2) PCT).

1.2 Claims 1-28 are not inventive (Article 33(3) PCT).

1.3 Claims, 1-28, are industrially applicable (Article 33(4) PCT).

**2.0 Novelty (Article 33(2) PCT).**

2.1 Document D1, which is considered to represent the most relevant state of the art, discloses the same problem (Page 431, Left Column, Last Paragraph - Right Column, First Paragraph) of finding synonymous information and discloses all the features of claim 1 (the references in parenthesis applying to this document):

- A method of processing a data record for finding a counterpart in a reference data set (Page 433, Left Column, Last Paragraph, "set of search terms". Note: A set of search terms is a data record and a database contains a reference data set);
- Determining in the data record a value of a data field, the data field representing an identifier (Page 433, Right Column, Third Paragraph, "query node". Note: A query node is a data field of the data record which represents an identifier);
- Determining from a set of predetermined identifier values at least one synonym candidate for the value of the data field (Page 433, Right Column, Third Paragraph, "synonym edges");

**INTERNATIONAL PRELIMINARY  
REPORT ON PATENTABILITY  
(SEPARATE SHEET)**

International application No.  
**PCT/FI2004/000331**

- Determining if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account writing variations, and if the predetermined synonym acceptance criterion is fulfilled, associating the value of the data field and the synonym candidate as synonymous (Page 433, Right Column, Third and Fourth Paragraphs, "threshold ... ontology". Note: Since the query terms are inside an ontology and an ontology takes into account different variations of arranging the terms, this determining step also takes into account writing variations as claimed);
- Searching for a counterpart for the data record by comparing to entries of the reference data set the value of the data field and/or synonym associated with the value of the data field after determining if the predetermined synonym acceptance criterion is fulfilled (Page 433, Right Column, Sixth Paragraph, "generate required queries". Note: This implies that the generated required queries contain the counterpart terms which are implicitly searched for within the database. Evidently, the search is performed after having determined that the terms were accepted based on the threshold).

2.1.1 The subject-matter of claim 1 is therefore not novel (Article 33(2) PCT).

2.2 Since Independent Claims 21, 26 and 28 are rewordings of the same features, the same objection applies mutatis mutandis. Therefore, claims 21, 26 and 28 are also not novel (Article 33(2) PCT).

2.4 Dependent claims 2 - 20, 22 - 25 and 27 are also disclosed in D1 and are thus also not novel (Article 33(2) PCT).

2.5 Therefore, Claims 1 - 28 are not new according to Article 33(2) PCT.

**3.0 Inventive Step (Article 33(3) PCT).**

3.1 Since Claims 1 - 28 are not new (cf, §2.5), claims 1 - 28 are therefore also not inventive (Article 33(3) PCT).

**4.0 Industrial Applicability (Article 33(4) PCT).**

**INTERNATIONAL PRELIMINARY  
REPORT ON PATENTABILITY  
(SEPARATE SHEET)**

International application No.  
**PCT/FI2004/000331**

4.1 Claims 1-28 fall within the technical field of Query Processing and are thus Industrially Applicable.

## Claims

1. A method of processing a data record for finding a counterpart in a reference data set, the method comprising the steps of:

5 determining in the data record a value of a data field, the data field representing an identifier,

determining from a set of predetermined identifier values at least one synonym candidate for the value of the data field,

10 determining if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account writing variations, and if the predetermined synonym acceptance criterion is fulfilled, associating the value of the data field and the synonym candidate as synonyms, and

15 searching for a counterpart for the data record by comparing to entries of the reference data set the value of the data field and/or a synonym associated with the value of the data field after determining if the predetermined synonym acceptance criterion is fulfilled.

2. A method as defined in claim 1, wherein the at least one synonym candidate is determined using a candidate selection criterion depending at least on the value of the 20 data field and on a synonym candidate.

3. A method as defined in claim 2, wherein the candidate selection criterion takes into account how similar a synonym candidate and the value of the data field sound.

25 4. A method as defined in claim 2, wherein the candidate selection criterion specifies that at least a predetermined part of the value of the data field is identical to a predetermined part of a synonym candidate.

30 5. A method as defined in any one of claims 2 to 4, wherein the candidate selection criterion takes into account also a further data field of the data record, said further data field representing a second identifier.

35 6. A method as defined in any preceding claim, wherein at least one quality parameter is evaluated for a synonym candidate, the synonym acceptance criterion taking into account the at least one quality parameter.

7. A method as defined in claim 6, wherein at least one quality parameter takes into account at least one of the following quantities:

a number of changes required for converting the value of the data field to be identical to a synonym candidate; a proportion of identical characters in the value of the data field and in a synonym candidate; and a difference between the length of the value of the data field and the length of a synonym candidate.

5

8. A method as defined in claim 7, wherein the number of changes required for converting the value of the data field to be identical to a synonym candidate is calculated using the Levenshtein distance.

10 9. A method as defined in claim 7, wherein the proportion of identical characters takes into account the order of the characters.

15 10. A method as defined in any one of claims 6 to 9, wherein a first quality parameter is evaluated for each synonym candidate and at least a second quality parameter is evaluated at least for the synonym candidate(s) having the best first quality parameter.

11. A method as defined in any one of claims 6 to 10, wherein the synonym acceptance criterion requires that there is only one synonym candidate having the best at least one quality parameter.

20

12. A method as defined in any one of claims 6 to 11, wherein at least two quality parameters are evaluated for each synonym candidate and the synonym candidate acceptance criterion specifies a threshold for one of the at least two quality parameters, the threshold being dependent on a further one of the at least two quality parameters.

25

13. A method as defined in any preceding claim, wherein the search for the counterpart involves comparison of the value of the data field to a synonym set relating to the identifier, members of said synonym set referring to respective predetermined identifier values, and when the predetermined synonym acceptance criterion is fulfilled, the value of the data field is added to the synonym set as a member referring to the synonym associated with the value of the data field before the search for the counterpart.

30 35 14. A method as defined in any preceding claim, wherein determining the at least one synonym candidate is discarded, if a predetermined discard criterion is fulfilled.

15. A method as defined in claim 14, wherein the predetermined discard criterion specifies that the value of the data field is identical to one of the predetermined identifier values.

5 16. A method as defined in claim 14, wherein the search for the counterpart involves the synonym set and the predetermined discard criterion specifies that the value of the data field is at least one of the following: one of the predetermined identifier values, and a member of the synonym set.

10 17. A method as defined in any one of claims 14 to 16, wherein the predetermined discard criterion takes into account a value of a second data field in the data record.

15 18. A method as defined in any preceding claim, wherein information indicating the at least one synonym associated with the value of the data field is added to the data record.

19. A method as defined in claim 18, wherein a copy of the data record is made for each synonym associated with the value of the data field.

20 20. A method as defined in any preceding claim, wherein the identifier relates to a name of one of the following: a geographical entity, a person and an organisation.

25 21. A method of processing a synonym set for searching counterparts in a reference data set for data records, a data record containing a data field representing an identifier, members of the synonym set being first identifier values and referring to respective second identifier values, the second identifier values being predetermined identifier values, and said searching for a counterpart involving comparison of a value of the data field to the synonym set, the method comprising the steps of determining among the predetermined identifier values at least one synonym candidate relating to the value of the data field in the data record, and, if the value of the data field and a synonym candidate fulfill a predetermined synonym acceptance criterion taking into account writing variations, adding before searching a counterpart for a data record the value of the data field to the synonym set as a member referring to the synonym candidate.

30 35 22. A method as defined in claim 21, wherein the synonym set is empty before adding the value of the data field to the synonym set.

23. A method as defined in claim 21, wherein the synonym set contains at least one member before adding the value of the data field to the synonym set.

24. A computer program comprising program instructions for causing a computer to 5 perform the method of any one of claims 1 to 23.

25. A computer program as defined in claim 24, embodied on a computer-readable record medium.

10 26. A data processing system for processing data records for finding counterparts in a reference data set, the system comprising:

- means for receiving data records,
- means for storing the reference data set,
- means for storing predetermined identifier values for an identifier,
- means for determining in the data records values of a data field, the data field representing the identifier,
- means for associating values of the data field and respective predetermined identifier values as synonyms, said means configured to determine from the predetermined identifier values at least one synonym candidate for a value of the data field, to determine if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account writing variations, and if the predetermined synonym acceptance criterion is fulfilled, to associate the value of the data field and the synonym candidate as synonyms, and
- means for searching counterparts in the reference data set for the data records, said searching involving comparing to entries of the reference data set values of data fields and/or synonyms associated with the values of the data fields.

20 27. A data processing system as defined in claim 26, further comprising  
25

- means for storing a synonym set, members of said synonym set referring to respective predetermined identifier values,
- wherein the means for associating values of the data field and respective predetermined identifier values as synonyms are configured to add to the synonym set a member referring to the synonym associated with the value of the data field before activation of the means for searching counterparts.

30 35 28. A data processing system for processing a synonym set for searching counterparts in a reference data set for data records, a data record comprising a data field representing an identifier, members of the synonym set being first identifier values

and referring to respective second identifier values, said second identifier values being predetermined identifier values, and said searching involving comparing a value of the data field to the synonym set, the system comprising:

- means for storing the synonym set,
- 5 - means for storing predetermined identifier values for the identifier,
- means for receiving data records,
- means for determining in the data records values of the data field, and
- means for adding to the synonym set a value of the data field and respective predetermined identifier values associated as synonyms before searching counterparts
- 10 in the reference data set, said means configured to determine from the predetermined identifier values at least one synonym candidate for a value of the data field, to determine if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account writing variations, and if the predetermined synonym acceptance criterion is fulfilled, to associate the
- 15 value of the data field and the synonym candidate as synonyms.